DOCUMENT RESUME

ED 131 128 TM 005 876

AUTHOR Willoughby, Lee; And Others
TITLE A Comparison of Domain-Referenced and Classic Psychometric Test Construction Methods.
PUB DATE 76
NOTE 13p.; Paper presented at the Annual Meeting of the Southeastern Psychological Association (New Orleans, Louisiana, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Criterion Referenced Tests; *Norm Referenced Tests; Statistical Analysis; *Test Construction; Test Reliability
IDENTIFIERS *Domain Referenced Tests

ABSTRACT

This study compared a domain referenced approach with a traditional psychometric approach in the construction of a test. Results of the December, 1975 Quarterly Profile Exam (QPE) administered to 400 examinees at a university were the source of data. The 400 item QPE is a five alternative multiple choice test of information a "safe" physician should know. Content of the exam covers the broad areas of Internal Medicine, Pediatrics, Obstetrics/Gynecology, Surgery, and Basic Science, as well as additional sub-topics. For purposes of this study, two 75 item tests were constructed by pulling from the 400 item QPE by two different strategies. The domain referenced approach was used to construct a 75 item test by a random sample of the 400 items. Selection of the 75 items with the highest point biserial item-total correlations represented the traditional psychometric approach to test construction. The exams were then rescored to obtain scores and item analysis data on the random and psychometric tests. Then, the two tests were compared with respect to distribution of p values (the proportion answering an item correctly), point biserial item-total correlations, student scores across medical school year level and reliability. The results were discussed with regard to their consistency with expectations of the domain referenced and psychometric approaches. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED131128

A Comparison of Domain - Referenced and Classic Psychometric Test Construction Methods[1]

Lee Willoughby

Martha Lea

Michael Whitley

University of Missouri - Kansas City

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Proponents of domain-referenced testing have emphasized the importance of a test accurately representing the domain which it represents. Development of the domain-referenced approach occurred in the context of the movement to increase specificity of educational objectives. Considering the merits of defining educational objectives over informal, sometimes ambiguous, objectives, it was logical to conclude that educational tests should accurately represent the educational objectives. In the domain-referenced approach one would define objectives and a corresponding domain of test items. Domain sampling has been described as more important than classic psychometric methods of test construction. The purpose of this study was to compare a domain-referenced approach with a traditional psychometric approach to the construction of tests.

METHOD

Data analyzed in this study were derived from the administration of the December, 1975 Quarterly Profile Exam (QPE) to

[1]A paper presented at the annual meeting of the Southeastern Psychological Association in New Orleans, 1976

TM005 876

400 examinees. The examinee group included 354 medical students, 3 physician faculty, 30 interns/residents, 8 health professionals and 5 non-faculty physicians. The number of Year I-VI students were 75, 82, 64, 71, 39, and 23, respectively. Slide 1 illustrates this data. The students are enrolled in a six year, combined B.S. and M.D. program at the University of Missouri - Kansas City.

The 400 item QPE is a 5 alternative, multiple choice test of information a "safe" physician should know. The QPE is one product of a computer-assisted test construction system functioning at the University of Missouri-Kansas City. This system in practice uses elements of domain-referenced and norm-referenced approaches. Content of the exam covers the broad areas of Internal Medicine, Pediatrics, Obstetrics/Gynecology, Surgery, and Basic Science, as well as additional sub-topics. For purposes of this study, two 75 item tests were constructed by pulling from the 400 item QPE by two different strategies. A random sample of the domain of 400 QPE items produced 75 unique items which would constitute the domain-referenced test. Selection of the 75 items with the highest point biserial item-total correlations represented the traditional psychometric approach to test construction. See slide 2.

The exams were then rescored to obtain scores and item analysis data for the domain-referenced and psychometric tests. Then, the two tests were compared with respect to score and item

characteristics.

RESULTS AND CONCLUSIONS

The mean and standard deviation of scores for the domain-referenced test were, in raw score units, 34.6 and 9.64, respectively; those figures for the psychometric test were 43.3 and 17.5[illegible], respectively. Mean performance was significantly ($t = 13.90$; $p < .01$) higher on the psychometric test; furthermore, score variance was significantly ($t = 29.19$; $p < .01$) greater on the psychometric test. However, the correlation of scores for the domain-referenced and psychometric tests was .904. The two scores correlated to a great extent, but differed with regard to central tendency and dispersion.

In the context that the QPE is used, the performance across Years I-VI is more important than the overall mean and standard deviation previously discussed. The exam is intended to evaluate the acquisition of information through six years of matriculation. The mean and range of performance on the domain-referenced and psychometric tests are presented in percent correct units by Year level in slide 3. The frequency distribution of domain-referenced and psychometric scores is presented by Year level in slide 4. That the psychometric approach yielded scores with greater variability was previously noted. The frequency distributions of the two tests indicate that the psychometric approach separate the scores by year level better than the domain-referenced

approach. Compared with the domain-referenced test, the psychometric test is more difficult at the lower level and easier at the upper level of the student body.

Now, consider the item characteristics of the two tests constructed. Slide 5 presents the frequency distribution of p values for the domain-referenced and psychometric tests. The distribution of p values is positively skewed with the domain-referenced test; the psychometric approach yielded a distribution of p values more closely approximating normality. Slide 6 presents the frequency distribution of item-total correlations for the domain-referenced and psychometric tests. Considering the manner in which the items were chosen for the psychometric test, the psychometric test was expected to have a more restricted range of item-total correlations than the domain-referenced test. It should be noted that the domain-referenced test contained a majority of items with significant item-total correlations. This, too, was expected since the 400 item QPE has a high level of internal consistency; the K-R formula 20 reliability coefficient was .954 for the QPE. Both the psychometric (.959) and domain-referenced (.845) reliability coefficients were quite respectable.

One final point of comparison remains to be reported. The question was asked, "How well is the content of the 400 item QPE represented by the domain-referenced and psychometric tests?" Slide 7 presents data which compares the content of these two tests with that of the total QPE. Recall, the

QPE covers five broad areas Internal Medicine, Pediatrics, Obstetrics/Gynecology, Surgery, and Basic Science. The proportional representation by content areas for the domain-referenced and psychometric tests did not significantly differ from that of the total QPE.

In summary, the results of two approaches to test construction differed with respect to central tendency, dispersion, and reliability. Also, the items of the two tests differed with respect to the distribution of p values and item-total correlations; they did not, however, differ with respect to the content of items compared with the total item pool.

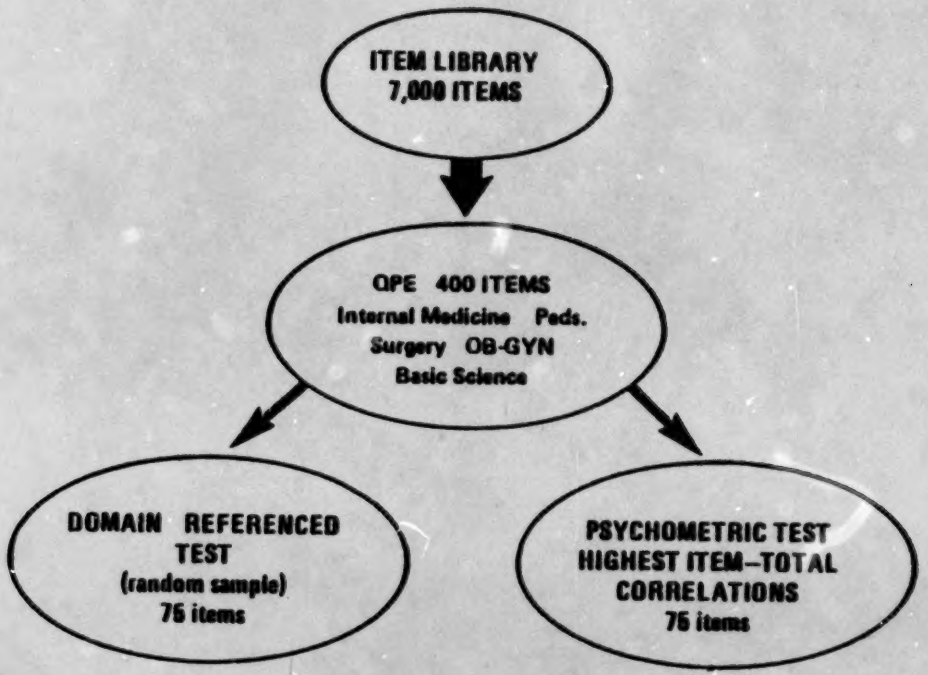
A couple more points should be made before concluding this paper. The domain referenced approach did yield a reliable measurement, although somewhat less reliable than the psychometric approach. This is probably due to the lengthy process of item generation and review involved in creating the QPE.

A second concluding point to consider involves the question of how fair to the domain-referenced concept is the random sample approach used in this study. A better approach might have been to use the D% (Year VI percent correct - Year I percent correct) to rank and select items. Well, that approach was attempted; however, the relationship of D% to item-total correlation was so high that the two approaches yielded 75 items each with 67 item common to both tests.

NUMBER OF SUBJECTS PER EXAMINEE CATEGORY

Six Year Combined B.S. and M.D. Program

Year I	75
Year II	82
Year III	64
Year IV	71
Year V	39
Year VI	23
Physician Faculty	3
House Staff	30
Health Professionals	8
Non-faculty Physician	5


ITEM LIBRARY
7,000 ITEMS
QPE 400 ITEMS
Internal Medicine Peds.
Surgery OB-GYN
Basic Science
DOMAIN REFERENCED
TEST
(random sample)
75 items
PSYCHOMETRIC TEST
HIGHEST ITEM–TOTAL
CORRELATIONS
75 items

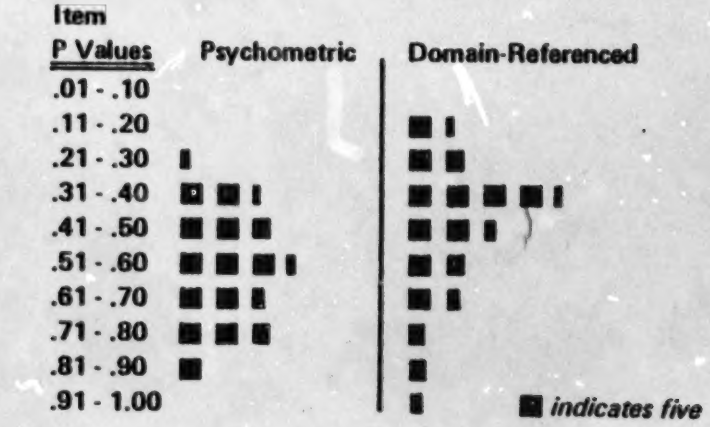
YEAR		MEAN	RANGE	
I	Psyc	24.7	12.0	37.3
(n = 75)	D-R	29.6	17.3	41.3
II	Psyc	41.4	21.3	76.0
(n = 82)	D-R	38.5	22.7	57.3
III	Psyc	61.6	41.3	86.7
(n = 64)	D-R	47.4	33.3	65.3
IV	Psyc	73.0	32.0	92.0
(n = 71)	D-R	53.4	45.3	72.0
V	Psyc	81.2	48.0	98.7
(n = 39)	D-R	58.2	38.7	78.7
VI	Psyc	86.1	73.3	98.7
(n = 23)	D-R	61.7	52.0	78.7

FREQUENCY DISTRIBUTION OF SCORES ACROSS YEAR LEVELS

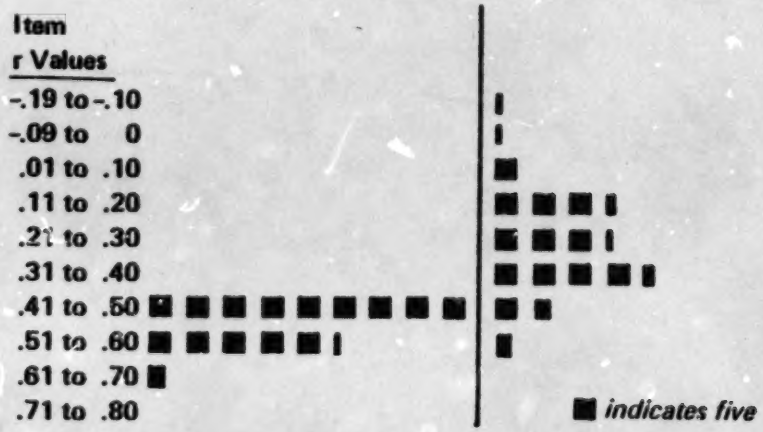
PER CENT CORRECT	Domain-Referenced Year						Psychometric Year					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI
11 - 15							3					
16 - 20	5						17					
21 - 25	12	3					[28]	3				
26 - 30	[25]	4					10	9				
31 - 35	21	19	3	3			13	9				
36 - 40	11	[26]	9	2	1		4	18				
41 - 45	1	18	21	10	1			[15]	5	1		
46 - 50		9	[10]	9	3			15	9	1	1	
51 - 55		2	11	[17]	6	5		8	5	3	0	
56 - 60		1	7	16	[14]	6		3	12	4	0	
61 - 65			3	10	8	[6]		0	[9]	5	0	
66 - 70				3	4	5		1	7	10	0	
71 - 75				1	1	0		0	11	[18]	9	2
76 - 80					1	1		1	5	16	9	2
81 - 85									0	4	[6]	6
86 - 90									1	5	7	[7]
91 - 95										4	6	5
96 - 100											1	1

[ ] *interval wherein class mean lies*

DISTRIBUTION OF P VALUES FOR PSYCHOMETRIC AND DOMAIN-REFERENCED TESTS



DISTRIBUTION OF ITEM-TOTAL CORRELATIONS FOR PSYCHOMETRIC AND DOMAIN-REFERENCED TESTS



PROPORTIONAL DISTRIBUTION OF ITEMS ACROSS CONTENT AREAS

	Psychometric	Domain - Referenced	Total QPE
Int. Med	.25	.17	.20
Peds	.21	.17	.19
Ob/Gyn	.16	.22	.18
Surg	.15	.14	.19
B S	.21	.24	.22